Open in app ↗ Sign up Sign in

Medium Search

How Researchers Queried 100Million WhatsApp Numbers Per Hour (And What Changed)

8 min read · Just now

Berend Watchus Follow

Listen Share

Author: Berend Watchus November 26, 2025. Cyber security blog article

WhatsApp
3 billion accounts, phone numbers, faces, all (reverse) searchable. Largest face, phone book. Not a Leak - WhatsApp Volunteered the Data
WhatsApp

“This is fine: A single university computer queries 100 million phone numbers per hour, and nobody at Meta notices.”

A recent academic study revealed that WhatsApp — the world’s largest messaging

platform with 3.5 billion users — had a massive privacy vulnerability that allowed unprecedented access to user data. Here's what happened, and why it matters.

Hey there! You are using WhatsApp: Enumerating Three Billion Accounts for Security and Privacy

WhatsApp, with 3.5 billion active accounts as of early 2025, is the world's largest instant messaging platform. Given...

arxiv.org

The Simple Explanation: What They Did Globally

Let's start with Netherlands as an example. Dutch mobile numbers follow a simple pattern:

- **Domestic format:** 06-XXXXXXXX (10 digits)
- **International format:** +31 6 XXXXXXXX

Researchers from the University of Vienna systematically generated and tested every possible Dutch mobile number combination:

+31 6 00000001 → "Is this on WhatsApp?" → No

+31 6 00000002 → "Is this on WhatsApp?" → Yes ✓

+31 6 00000003 → "Is this on WhatsApp?" → No

...and so on for all 100 million possibilities

At 7,000 queries per second, they scanned the entire Dutch number space in under 4 hours. They found 18.7 million active accounts — more than the country's entire population.

Then they did this for 245 countries worldwide.

BTW: Why More Accounts Than People?

Netherlands has ~18 million people, but researchers found 18.7 million active WhatsApp accounts (103.6% of the population).

This happens because:

Multiple SIM cards per person: Many people have:

- A personal phone number
- A work phone number
- An old number they still keep active
- Separate numbers for business/privacy

Non-residents with Dutch numbers:

- Expats who moved away but kept their Dutch number
- International workers/students temporarily in Netherlands
- Business travelers with local numbers

Businesses and services:

- Companies running WhatsApp Business accounts
- Customer service lines
- Delivery services, etc.

Inactive population counted:

- Young children (under 13) who don't use WhatsApp
- Elderly citizens without smartphones
- But their "share" of numbers is used by others with multiple accounts

The key point: When you see "103% per capita," it doesn't mean WhatsApp made a mistake — it reveals that the phone number system is saturated, with heavy users claiming multiple slots

The Scale of the Problem

Using a tool they built called "libphonegen," the researchers:

- Generated 63 billion plausible mobile phone numbers globally
- Queried WhatsApp's servers from a single university computer
- Encountered **no rate limiting or blocking whatsoever**
- Discovered **3.5 billion active WhatsApp accounts** between December 2024 and April 2025

For context: that's more data than the massive 2013 Yahoo breach (3 billion accounts) and five times larger than the 2021 Facebook scraping incident (500 million records).

What Data Was Exposed

For each discovered account, they could retrieve:

- **Phone number** (obviously)
- **Profile pictures** (if set to public — they downloaded 77 million; 66% contained recognizable faces)
- **About/status text** (revealing personal details, political views, even drug dealers advertising prices)
- **Device information** (Android vs. iOS — useful for targeted attacks)
- **Encryption keys and timestamps** (revealing device age, account activity, companion devices)
- **Business account information**

The Most Alarming Findings

1. Leaked Data Never Dies Comparing their dataset with the 2021 Facebook leak: **58% of those phone numbers are still active on WhatsApp six years later.** This

means once your number is leaked, it remains a target for scams and phishing indefinitely.

2. Users in Banned Countries at Risk They found active accounts in countries where WhatsApp is officially banned:

- China: 2.3 million accounts
- Myanmar: 1.6 million accounts
- Iran: 60+ million accounts (before ban was lifted)
- North Korea: 5 accounts

Using WhatsApp in some of these countries carries serious legal consequences.

Get Berend Watchus's stories in your inbox

Join Medium for free to get updates from this writer.

Enter your email

Subscribe

3. Broken Encryption Keys They discovered massive security issues in WhatsApp's cryptographic implementation:

- 2.3 million public encryption keys were reused across different devices
- Some keys appeared hundreds of thousands of times
- 20 accounts (mainly in the US) used an **all-zero private key** — essentially no encryption at all

4. The "Reverse Phone Book" Threat With 77 million profile pictures downloaded, facial recognition could create a searchable database: upload a photo, get a phone number. Add leaked Facebook data (names, birthdates, employers), and you have a

'stalker's/criminal paradise'.

The Multi-Directional Threat: A Database That Works Every Way

What makes this dataset uniquely dangerous isn't just its size — it's its **versatility.** Unlike previous leaks that work in one direction, this creates a web of connections exploitable from multiple entry points:

Every Search Direction Works

Phone Number → Identity

- Input: +1–555–0123
- Output: Face, name, status, location (from area code), device type

Face → Phone Number (Reverse Image Search)

- Upload a photo from dating apps, social media, or security cameras
- Match against the facial database
- Get their direct phone number

Name → Phone Number

- Search "John Smith, Netherlands"
- Cross-reference with area codes and profile names/status text
- Find specific individuals among common names

Location → Target List

- Filter by country/area code
- Generate lists like "all WhatsApp users in Myanmar" (regime targeting) or "all users in wealthy zip codes" (scam targeting)

Device Type → Exploit Targeting

- Filter for iOS users (perceived wealth)
- Target Android users with Android-specific malware
- Identify outdated devices via encryption key analysis

The Enrichment Multiplier

When combined with existing breaches, each data point validates and enriches the others:

- **2021 Facebook leak** (names, birthdays, employers) + WhatsApp data = verified current contact info with faces
- **LinkedIn scrapes** (professional history) + WhatsApp = corporate espionage targeting
- **Dating app photos** + WhatsApp facial database = identify and contact people who ghosted you (stalking)
- **CCTV/public photos** + facial matching = identify strangers and get their phone numbers

Why This Is Unprecedented

Previous major leaks worked in limited directions:

- Facebook 2021: Phone → Profile data (but no systematic facial database)
- Yahoo 2013: Email → Account data (but no phones or faces)
- LinkedIn 2021: Name → Professional data (but no verified contact methods)

This WhatsApp dataset enables: ✅ Bidirectional searches (any field finds all others)
✅ Biometric identification (faces)
✅ Direct contact capability (phone numbers)
✅ Real-time verification (active as of 2025)
✅ Global geographic precision (area codes for 245 countries)

✅ Cross-platform enrichment (linkable to all previous leaks)

🚨 **The reality:** This is the largest **multi-indexed, biometric-enabled contact database** ever compiled. It's not just a phone book with faces — it's a surveillance tool that works from any angle: face, name, location, or number. All roads lead to complete identification.

In the wrong hands, this doesn't just enable targeted attacks — it enables **mass surveillance at a scale previously only available to nation-states.**

⚠️ Why WhatsApp Didn't Stop Them

The vulnerability stems from WhatsApp's core feature: **contact discovery**. When you install WhatsApp, it checks which contacts from your address book are on the platform. This requires the ability to query phone numbers.

The problem? WhatsApp assumed rate limiting would prevent abuse. It didn't.

Here's what makes this truly alarming: The researchers made **zero effort to hide their activity:**

- Used the same university IP address for four months straight (no IP rotation)
- Used only five WhatsApp accounts (no account cycling)
- Their server had a publicly listed abuse email contact
- Sent queries at a blistering 7,000 per second, 24/7
- Were identifiable as an academic institution via reverse DNS lookup

Nobody at WhatsApp noticed. Nobody reached out. No automated systems triggered.

For a platform handling 3.5 billion users and owned by Meta — a company that should have sophisticated monitoring systems — this represents a staggering oversight. Imagine if a malicious actor had been doing this instead of academic researchers operating in good faith.

🚨 The researchers were essentially waving a giant red flag labeled "UNIVERSITY RESEARCH PROJECT" while downloading data on half the planet's population. **And WhatsApp's detection systems saw nothing wrong.**

🚨🚨 **The darker implication:** If academic infrastructure provides such effective **cover** for mass data extraction, what's to stop malicious actors from routing their scraping operations through university networks or disguising themselves as researchers? This study may have inadvertently revealed not just a vulnerability, but a **blueprint for how to exploit it undetected.**

The Responsible Disclosure Saga

This is where it gets frustrating:

September 2024: First report to Meta about lack of rate limiting
September 24, 2024: Meta acknowledges, says they'll forward to engineering
March 28, 2025: Second report explicitly mentioning phone enumeration risk
April 14, 2025: Third report revealing they'd successfully enumerated 3.5 billion accounts
April 17, 2025: Meta finally responds: "We're reviewing"
August 2025: Researchers threaten to publish; Meta requests delay
September-October 2025: Active collaboration begins
October 2025: Issues finally fixed

What Changed: The Fixes

WhatsApp implemented several countermeasures (confirmed effective by October 2025):

1. Smart Rate Limiting

- New "cardinality counters" track how many *unique* accounts you query over your account's lifetime
- Machine learning identifies scraping patterns early
- Limits apply without breaking legitimate contact discovery

2. Profile Data Restrictions

- Even if you set profile picture to “Everyone,” there are now limits on bulk access
- About text visibility similarly restricted
- Exception: Business accounts remain public (by design)

3. Timestamp Removal

- Profile picture retrieval no longer includes “last changed” timestamps

4. Key Reuse Bug Fixed

- Android client bug causing key reuse during phone number changes patched (version 2.25.26.11+)

What This Means for You

Immediate Actions:

1. **Review privacy settings:** Set profile photo and about text to “Contacts Only” (this is now default for new users)
2. **Assume your number is known:** If you had WhatsApp before October 2025, assume your phone number’s WhatsApp status is compromised
3. **Be wary of phishing:** Attackers may cross-reference this data with old leaks (Facebook 2021, etc.)

The Bigger Picture: This study reveals a fundamental tension in messaging apps: convenience (easy contact discovery) vs. privacy. Even with end-to-end encryption protecting message *content*, massive amounts of *metadata* remain exposed.

Alternative solutions exist — Signal uses privacy-preserving contact discovery, and encrypted profile information — but adoption is limited.

Why This Research Matters

Some might ask: “Isn’t this irresponsible disclosure?” The researchers argue the opposite. As previous studies have shown, **empirical testing is the only way to verify real-world security.** Without actually attempting enumeration at scale, WhatsApp might never have prioritized fixing these issues.

The data was handled responsibly:

- All personal data was deleted before publication
- No individual users were analyzed
- Only aggregate statistics were published
- WhatsApp confirmed fixes before the paper’s release

The Bottom Line

For over a year, anyone with basic technical skills could have built a database of 3.5 billion WhatsApp users, complete with photos and metadata. While we can’t know if malicious actors exploited this, the researchers demonstrated it was trivially easy.

The good news: it’s now fixed. The bad news: your phone number’s WhatsApp status and any public data from before October 2025 may already be in unknown hands.

WhatsApp’s final statement acknowledged the issue and thanked the researchers for their collaboration. But it took nearly a year and multiple escalations to get there — a reminder that even billion-user platforms can overlook critical vulnerabilities until researchers force the issue.

Read the full paper: arXiv:2511.20252 Nov 26, 2025 on arXiv.org
Authors: Gabriel K. Gegenhuber, Philipp É. Frenzel, Maximilian Günther, Johanna Ullrich, Aljosha Judmayer (University of Vienna, SBA Research)

Cybersecurity | Cyber Security Awareness | Infosec | Telco | Information Security

Follow

Written by Berend Watchus

23 followers · 45 following

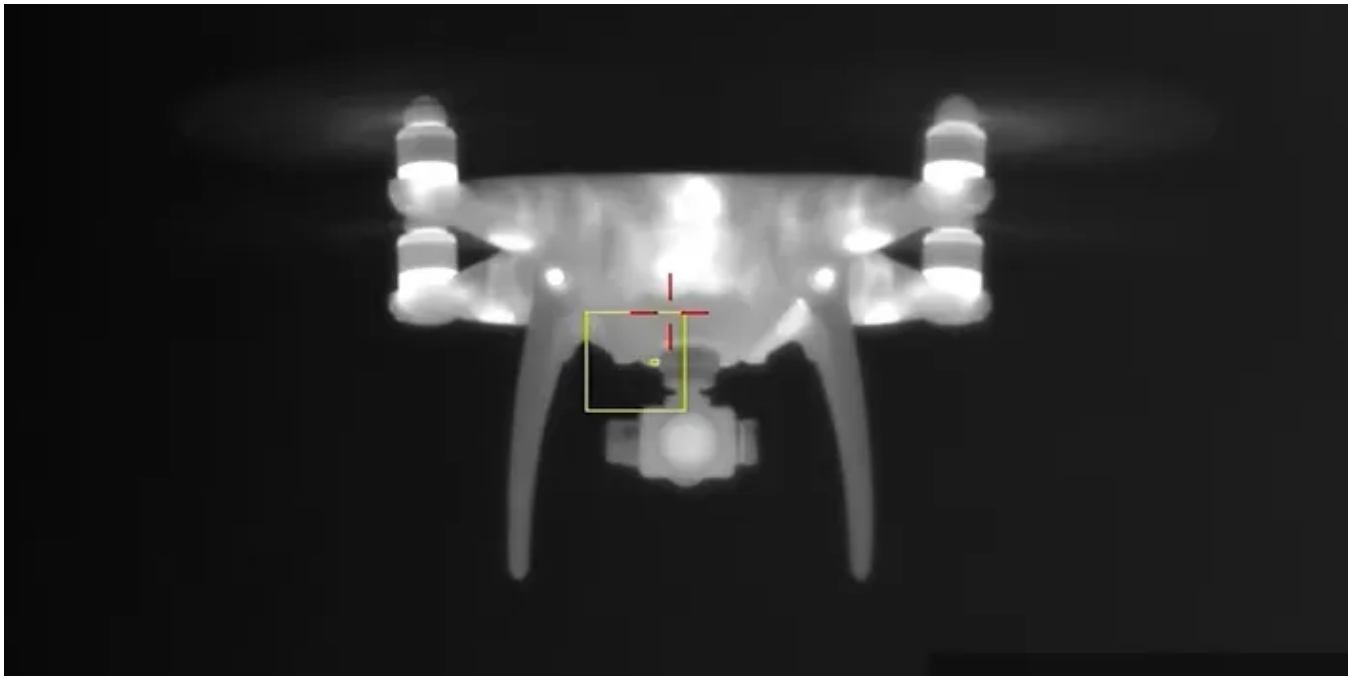
Independent Researcher (Non-Profit) exploring AI & interdisciplinary solutions. Former futurist event organizer & care innovation lab volunteer

No responses yet

Write a response

What are your thoughts?

More from Berend Watchus



In System Weakness by Berend Watchus

🔐 NATO Counter-Drone Operations in Belgium: Deep Technical Analysis for Cyber Security...

Author: Berend Watchus

Nov 12

In System Weakness by Berend Watchus

Linguistic Trojan Horse: Why LLM Steganography Just Broke AI Safety

Author: Berend Watchus October 24, 2025 Blog article

Oct 24 1

In System Weakness by Berend Watchus

Whisper Leak: Your LLM Chats Are Not as Private as You Think (Even Encrypted!)

Author: Berend Watchus November 11, 2025 Blog article

Nov 7

Berend Watchus

The $20,000 Robot That Isn’t Really a Robot: Inside the Neo Deception

Author: Berend Watchus October 30, 2025 Blog article

Oct 30

See all from Berend Watchus

Recommended from Medium

Shahzaib READ NOW !

In ILLUMINATION by Shahzaib

I Used WormGPT for Free!

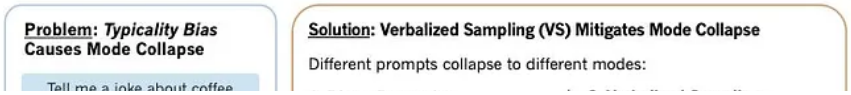
What Happened When I Tested the Internet's Most Dangerous AI Tool

Nov 18 356 7

:2510.01171v3 [cs.CL] 10 Oct 2025

ABSTRACT

Post-training alignment often reduces LLM diversity, leading to a phenomenon known as *mode collapse*. Unlike prior work that attributes this effect to algorithmic limitations, we identify a fundamental, pervasive data-level driver: *typicality bias* in preference data, whereby annotators systematically favor familiar text as a result of well-established findings in cognitive psychology. We formalize this bias theoretically, verify it on preference datasets empirically, and show that it plays a central role in mode collapse. Motivated by this analysis, we introduce ***Verbalized Sampling (VS)***, a simple, training-free prompting strategy to circumvent mode collapse. VS prompts the model to verbalize a probability distribution over a set of responses (e.g., "Generate 5 jokes about coffee and their corresponding probabilities"). Comprehensive experiments show that VS significantly improves performance across creative writing (poems, stories, jokes), dialogue simulation, open-ended QA, and synthetic data generation, without sacrificing factual accuracy and safety. For instance, in creative writing, VS increases diversity by 1.6-2.1× over direct prompting. We further observe an emergent trend that more capable models benefit more from VS. In sum, our work provides a new data-centric perspective on mode collapse and a practical inference-time remedy that helps unlock pre-trained generative diversity.


Problem: Typicality Bias Causes Mode Collapse
Solution: Verbalized Sampling (VS) Mitigates Mode Collapse
Different prompts collapse to different modes:

In Generative AI by Adham Khaled

Stanford Just Killed Prompt Engineering With 8 Words (And I Can't Believe It Worked)

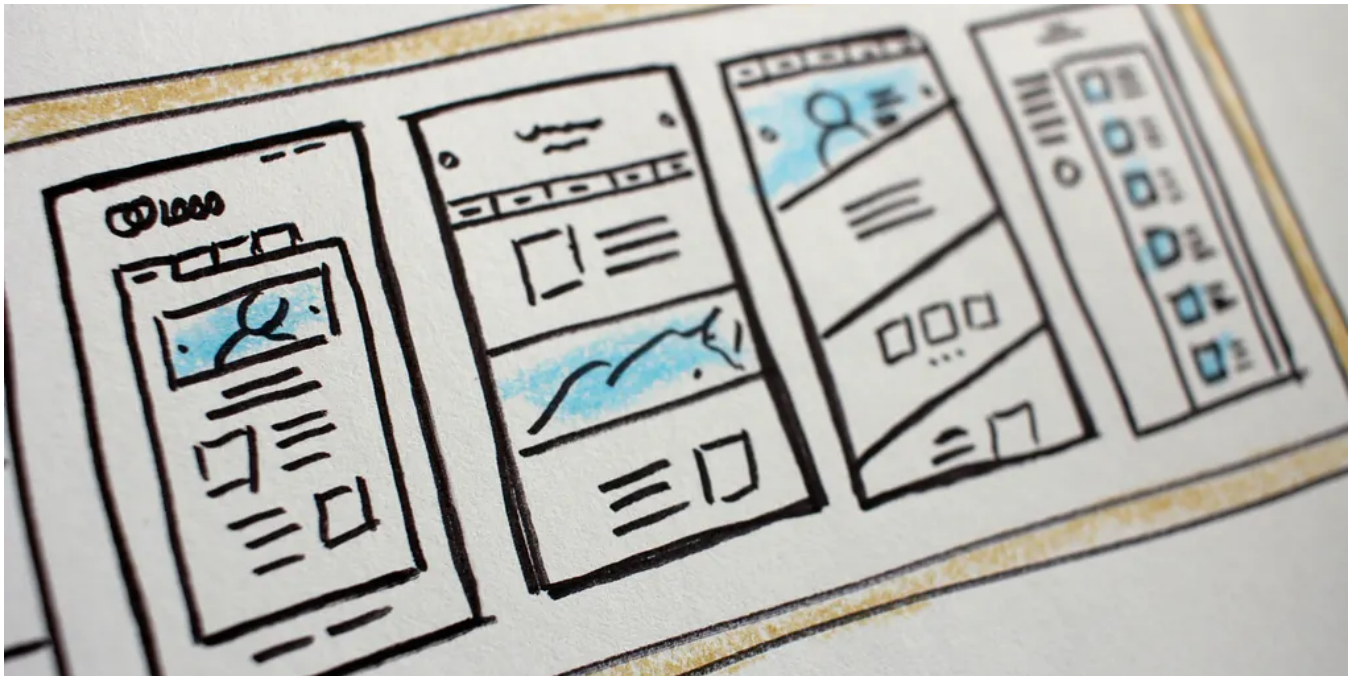
ChatGPT keeps giving you the same boring response? This new technique unlocks 2× more creativity from ANY AI model — no training required...

Taimur Ijlal

The Cybersecurity Jobs That AI Will Kill — And the Ones That Will Survive

AI Is Reshaping Cybersecurity Careers — Here's What Survives (and What Doesn't)

6d ago 80 1



Burk

The End of Personal Websites

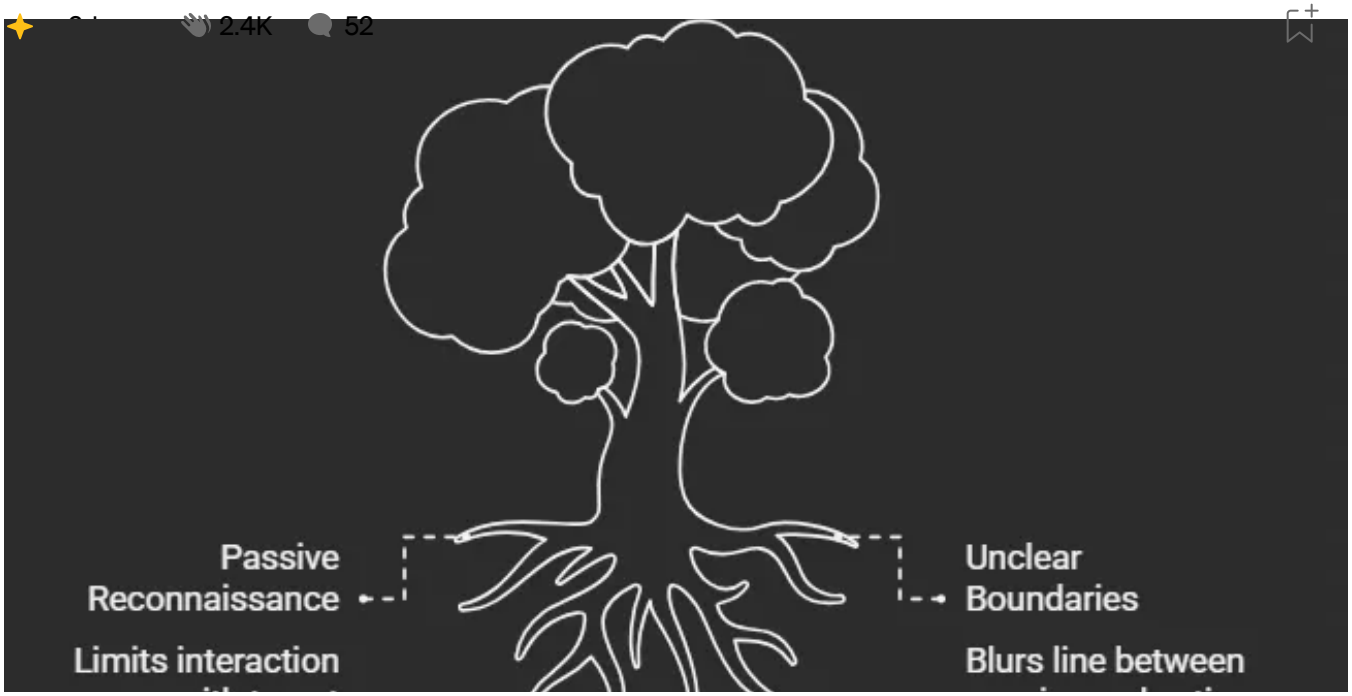
And why that might be a good thing

Nov 16 6.8K 262

Will Lockett

Did Nvidia Just Prove There Is No AI Bubble?

Take a wild guess.



Mainekhacker

OSINT detailed guide for 2026

1.0 Introduction

Nov 17 18

See more recommendations